542043

P1 1153916

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

April 14, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/441,046*
FILING DATE: *January 17, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/01329*

REC'D 16 APR 2004
WIPO PCT

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. R. Grant

P. R. GRANT
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

01-21-03604U1046 .01170304

Approved for use through 10/31/2002. OMB 0651-0032
PTO/SB/16 (10-01)
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EL565095584US

INVENTOR(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Charles R.	Cantor	Del Mar, CA
Chunming	Ding	Waltham, MA

☐ Additional inventors are being named on the _____ separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max)

HAPLOTYPE ANALYSIS

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☒ Customer Number 26,248 → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm or Individual Name	Ronald I. Eisenstein/Leena H. Karttunen				
Address	NIXON PEABODY LLP				
Address	101 Federal Street				
City	Boston	State	MA	ZIP	02110
Country	US	Telephone	(617) 345-6054/1367	Fax	(866) 743-2115/(617) 345-1300

ENCLOSED APPLICATION PARTS (check all that apply)

☒ Specification *Number of Pages* 15
☒ Drawing(s) *Number of Sheets* 3
☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number
☒ Other (specify) Fee Transmittal; Express Mail Cert; Check; Return Receipt Postcard

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☒ Applicant claims small entity status. See 37 CFR 1.27.
☒ A check or money order is enclosed to cover the filing fees
☒ The Commissioner is hereby authorized to charge ~~filing fees~~ or credit any overpayment to Deposit Account Number: 50-0850 - charge deficiencies
☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): 80.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.
☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE _____ Date 1/17/2003

TYPED or PRINTED NAME Ronald I. Eisenstein/Leena H. Karttunen

REGISTRATION NO. *(if appropriate)* 30,628/Lim. Rec. 37 CFR§10.9(b)

TELEPHONE (617) 345-6054 / 1367

Docket Number: 701586-053650-P

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is used by the public to file (and by the PTO to process) a provisional application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the complete provisional application to the PTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S.D Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

PTO/SB/17 (01-03)
Approved for use through 04/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

FEE TRANSMITTAL for FY 2003

Effective 01/01/2003. Patent fees are subject to annual revision.

[X] Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT ($) 80.00

Complete if Known	
Application Number	To be assigned
Filing Date	To be assigned
First Named Inventor	Cantor et al.
Examiner Name	To be assigned
Art Unit	To be assigned
Attorney Docket No.	701586-053650-P

METHOD OF PAYMENT *(check all that apply)*

[X] Check [ ] Credit card [ ] Money Order [ ] Other [ ] None

[X] Deposit Account:

Deposit Account Number: 50-0850 - charge deficiencies

Deposit Account Name: NIXON PEABODY LLP

The Commissioner is authorized to: *(check all that apply)*

[ ] Charge fee(s) indicated below [ ] Credit any overpayments

[ ] Charge any additional fee(s) during the pendency of this application

[ ] Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	80.00

SUBTOTAL (1) ($) 80.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

		Extra Claims		Fee from below	Fee Paid
Total Claims		-20** =		X	=
Independent Claims		-3** =		X	=
Multiple Dependent					=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($)

**or number previously paid, if greater; For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet	
1053	130	1053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1,450	2254	725	Extension for reply within fourth month	
1255	1,970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,300	2453	650	Petition to revive - unintentional	
1501	1,300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($)

SUBMITTED BY

Name (Print/Type)	Ronald I. Eisenstein/Leena H. Karttunen	Registration No. (Attorney/Agent)	30,628/37CFR§10.9(b)	(Complete *(if applicable)*) Telephone	(617) 345-6054/1367
Signature				Date	1/17/2003

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

If you need assistance completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

60044104-6 . 011703

Docket No. 701586-053650-P ***PATENT***

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Charles R. Cantor and Chunming Ding
Application No.: To be assigned
Filed: To be assigned
Title: HAPLOTYPE ANALYSIS

BOX PROVISIONAL APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231

EXPRESS MAIL CERTIFICATE

"Express Mail" Label Number: EL565095584US
Date of Deposit: January 17, 2003

I hereby state that the following *attached* papers and fees:

1. Provisional Patent Application (18 pp.): Specification – 14 pp., Claims – 1 pg., and Drawings 3 pp.;
2. Provisional Application for Patent Cover Sheet (1 pg);
3. Fee Transmittal (1 pg.);
4. Check in the Amount of $ 80.00;
5. Limited Recognition Statement for Leena H. Karttunen; and
6. Return Receipt Postcard

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. section 1.10, on the date indicated above and is addressed to Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Nicole Gignac

Nicole M. Gignac

Signature of person mailing paper or fee

BEFORE THE OFFICE OF ENROLLMENT AND DISCIPLINE
UNITED STATE PATENT AND TRADEMARK OFFICE

LIMITED RECOGNITION UNDER 37 CFR § 10.9(b)

Leena Karttunen is hereby given limited recognition under 37 CFR § 10.9(b), as an employee of the law firm of Nixon Peabody LLP, to prepare and prosecute patent applications wherein the patent applicant is a client of the law firm of Nixon Peabody LLP, and a registered practitioner, who is a member of the law firm of Nixon Peabody LLP, is the practitioner of record in the applications. This limited recognition shall expire on the date appearing below, or when whichever of the following events first occurs prior to the date appearing below: (i) Leena Karttunen ceases to lawfully reside in the United States, (ii) Leena Karttunen's employment with the law firm of Nixon Peabody LLP, ceases or is terminated, or (iii) Leena Karttunen ceases to remain or reside in the United States on an H-1B visa.

This document constitutes proof of such limited recognition. The original of this document is on file in the Office of Enrollment and Discipline of the U.S. Patent and Trademark Office.

Expires: June 14, 2003

Harry I. Moatz
Director of Enrollment and Discipline

Attorney Docket No. 701586-53650-P
Express Mail Label No. EL565095584US

HAPLOTYPE ANALYSIS

BACKGROUND OF THE INVENTION

[001] Single nucleotide polymorphism (SNP) analysis has drawn much attention with the hope of identifying genetic markers for common diseases. However, current genotyping technologies are only able to determine each SNP separately. In other words, there is a lack of information on how several SNPs are associated with each other physically on a chromosome. A DNA haplotype, the phase determined association of several DNA markers (e.g. SNPs) is a much more powerful method statistically for disease association studies. Yet unfortunately, it is also much harder to determine a haplotype. Current experimental approaches include a physical separation of homologous chromosomes via means of mouse cell line hybrid, cloning into a plasmid and allele specific PCR. Neither of them is simple enough a method for routine high-throughput analysis. There are also ways to computationally determine haplotypes, but the accuracy of such computational analysis is uncertain.

[002] Haplotypes, combinations of several phase-determined polymorphic markers in a chromosome, are extremely valuable for studies like disease association [1,2] and chromosome evolution. Direct molecular haplotyping has relied heavily on family data, but is limited to short genomic regions (a few kilobases). Statistical estimation of haplotype frequencies can be inconclusive and inaccurate [3]. It would be desirable to have a method that enables direct molecular haplotyping over longer genomic regions.

[003] With the rapid discovery and validation of several million single nucleotide polymorphisms (SNP), it is now increasingly practical to use genome-wide scanning to find genes associated with common diseases [1,2]. However, individual SNPs have limited statistical power for locating disease susceptibility genes. Haplotypes, combinations of several phase-determined polymorphic markers, can provide additional statistical power in the mapping of disease genes [4-7].

[004] Haplotype determination of several markers for a diploid cell is complicated since conventional genotyping techniques cannot determine the phases of several

different markers. For example, a genomic region with three heterozygous markers can yield 8 possible haplotypes. This ambiguity can, in some cases, be solved if pedigree genotypes are available. However, even for a haplotype of only 3 markers, genotypes of father-mother-offspring trios can fail to yield offspring haplotypes up to 24% of the time. Computational algorithms such as expectation-maximization (EM), subtraction and PHASE are used for statistical estimation of haplotypes [4,8,9]. However, these computational methods have serious limitations in accuracy, number of markers and genomic DNA length. For example, for a haplotype of only 3 markers from doubly heterozygous individuals, the error rates of the EM and PHASE methods for haplotype reconstruction can be as high as 27% and 19%, respectively [3]. Alternatively, direct molecular haplotyping can be used based on the physical separation of two homologous genomic DNAs prior to genotyping. DNA cloning, somatic cell hybrid construction, allele specific PCR and single molecule PCR [10-12] have been used, and these approaches are largely independent of pedigree information. These methods are limited to short genomic regions (allele-specific PCR and single molecule PCR) and are prone to errors.

SUMMARY OF THE INVENTION

[005] The present invention provides an efficient way for high throughput haplotype analysis. Several SNP markers can be simultaneously determined through multiplex PCR of single DNA molecules. These SNPs can be of any distance to each other on the chromosome, something almost impossible to achieve for other methods. In addition, our overlapping SNP approach can be used to link smaller haplotypes into larger haplotypes. The invention provides a powerful new tool for diagnostic haplotyping.

[006] The method of the present invention enables direct molecular haplotyping of several polymorphic markers separated by as far as about 25 kilobases (kb). In short, genomic DNA is first diluted to approximately single copy. The haplotype is subsequently determined by genotyping several polymorphic markers in the same reaction with multiplexed PCR and base extension reactions. This approach does not rely on pedigree data and does not require previous amplification of the genomic region containing the selected markers.

[007] In one embodiment, the invention provides a method of haplotyping comprising the steps of identifying at least two SNPs within a genomic region; obtaining

a DNA sample; amplifying the DNA sample comprising more than one copy of the genomic DNA with at least two primer pairs each capable of amplifying a different region flanking each of the SNPs in separate reactions and individually genotyping the amplified DNA; diluting the DNA sample so that substantially only one copy of a genomic DNA is present in the dilution; amplifying the diluted DNA sample using the same primer pairs and performing a multiplex genotyping of the amplified DNA; determining the haplotype of the one copy of the genomic DNA genotyping the amplified sample from the single genomic DNA copy dilution. In a preferred method the genotyping is performed using primer expension, terminator nucleotides and MALDI-TOF analysis.

[008] In another embodiment, the invention provides a method of haplotyping comprising obtaining a genomic DNA sample; diluting the DNA sample so that substantially only one copy of a genomic DNA is present in the dilution; identifying at least two SNPs within a genomic region; amplifying the diluted DNA sample using primer pairs designed to amplify a genomic region flanking the SNPs; and determining the haplotype of the copy of the genomic DNA by multiplex genotyping.

[009] In a preferred embodiment, the methods of the present invention use mass spectrometry, for example MassArray system, to genotype the samples.

[010] Preferably at least two, more preferably at least three, four or five primer pairs are used simultaneously to amplify the polymorphic genomic regions. In one embodiment, at least ten primer pairs are used in the same reaction. In one embodiment, the haplotype is determined using at least two different combinations of primer pairs in at least two parallel reactions.

BRIEF DESCRIPTION OF FIGURES

[011] Figures 1A-1B show a flow chart of multiplex genotyping of single DNA molecules for haplotype analysis. Traditional genotyping methods using a few nano-grams (ng) genomic DNA (about 1600 copies of genomic templates) yield only the genotypes of each individual SNP marker, but the phases of these SNPs are not determined (shown in top right in the mass spectra in Fig. 1A). Simultaneous genotyping of several markers using multiplex assays with single DNA molecules (Fig. 1B) allows haplotyping analysis since the two alleles can be physically separated with very dilute

DNA concentrations (shown in bottom right in the mass spectra in Fig. 1B). In contrast to other molecular haplotyping methods, the entire haplotype block is not amplified in this approach. Instead, only about 100 bp around each individual SNP is amplified for genotyping, resulting in very high efficiency of PCR amplification from single DNA molecules. The SNP markers can be as far apart as desired, as long as there is no significant break between them.

[012] Figure 2 shows effects of genomic DNA concentration on haplotyping efficiency. About 3 pg, 5 pg and 9 pg (or 1, 1.6 and 3 copies of genomic templates) were used for haplotyping of 3 SNP markers in the CETP region. The DNA copy number in a specific reaction is estimated by the Poisson distribution. The haplotyping result can either be a failed assay, successful haplotyping, both alleles present (no phase determination for the markers), or an incomplete multiplex. Except for incomplete multiplexes, values are percentages from 54 to 144 individual multiplex assays (see methods section for details on the calculation), followed by predicted values using the Poisson distribution.

[013] Figure 3 shows overlapping multiplex genotyping assays with single DNA molecules. Seven SNP markers (A: rs289744, B: rs2228667, C: rs5882, D: rs5880, E: rs5881, F: rs291044, G: 2033254) from an 8kb genomic region of the CETP locus were chosen (details of these SNPs, their chromosome position and oligonucleotides used for genotyping are provided in Table 2). Two 5-plex genotyping assays were designed for these 7 markers and the overlapping heterozygous SNPs were used to obtain the entire haplotype of 7 SNP markers. Assays on individual 6 were used to demonstrate how this is carried out. Multiplex assay 1 determined the haplotype of 5 SNPs as AGAGT and CGGGC. Multiplex assay 2 determined the other haplotype of 5 SNPs as GGGCT and AGGTT. Then, the genotypes of the overlapping SNPs (SNP C, E, F) were used to combine the two 5-SNP haplotypes into a haplotype of 7 SNPs covering the entire region under investigation.

DETAILED DESCRIPTION OF THE INVENTION

[014] The present invention provides a direct molecule haplotyping approach which is based upon a surprising discovery that a single molecule dilution of genomic DNAcan be used for separation of two homologous genomic DNAs. The diluted sample is used in

direct multiplex genotyping of several markers with, for example, the MassArray™ system (Fig. 1). The approach of the present invention differs significantly from previous single molecule PCR method in that the method of the present invention does not require the amplification of the complete genomic region containing the markers of interest; thus it is not limited to only a few kb DNA. Close to 100% genotype success rates for single DNA molecules has been achieved. Additionally, the multiplex genotyping assay approach enables direct haplotype determination without pedigree genotype information. High throughput haplotyping can easily be achieved by incorporating the method of the present invention with any commercially available genotyping systems, such as the MassArray™ system.

[015] The invention provides a method of creating haplotypes comprising the steps of obtaining a DNA sample; amplifying the DNA sample comprising more than one copy of the genomic DNA with at least two primer pairs designed to amplify a genomic region containing a nucleic acid polymorphism and genotyping the amplified DNA; diluting the DNA sample so that substantially only one copy of a genomic DNA is present in the dilution; amplifying the diluted DNA sample using at least two primer pairs designed to amplify a genomic region containing a nucleic acid polymorphism and genotyping the amplified DNA; determining the haplotype of the one copy of the genomic DNA.

[016] **Genomic DNA.** Genomic DNA can be isolated from a subject using any method of DNA isolation known to one skilled in the art. Examples of DNA isolation methods can be found in general laboratory manuals, such as Sambrook and Russel, MOLECULAR CLONING: A LABORATORY MANUAL, 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001), the entirety of which is herein incorporated by reference

[017] **Polymorphic Markers and Oligonucleotides.** The number of polymorphic nucleic acid useful according to the present invention is ever increasing. Currently, such markers are readily available from a variety of publicly accessible databases and new ones are constantly being added to the pool of available markers. Particularly useful markers according to the present invention are single nucleotide polymorphisms (SNP). Examples of SNP databases include, but are not limited to Human SNP Database at http://www-genome.wi.mit.edu/snp/human, NCBI dbSNP Home Page at

http://www.ncbi.nlm.nih.gov/SNP, http://lifesciences.perkinelmer.com/SNPDatabase/welcome.asp, Celera Human SNP database at http://www.celera.com/genomics/academic/home.cfm?ppage=cds&cpage=snps, the SNP Database of the Genome Analysis Group (GAN) at http://www-gan.iarc.fr/SNPdatabase.html,

[018] A number of nucleic acid primers are already available to amplify DNA fragments containing the polymorphisms and their sequences can be obtained, for example, from the above-identified databases. Additional primers can also be designed, for example, using a method similar to that published by Vieux, E.F., Kwok, P-Y and Miller, R. D. in BioTechniques (June 2002) Vol. 32. Supplement: "SNPs: Discovery of Marker Disease, pp. 28-32.

[019] **Genotyping.** A number of different single nucleotide polymorphism (SNP) genotyping methods useful according to the present invention are known to one skilled in the art. Generally, SNP genotyping is performed after amplification of the genomic DNA samples using, for example polymerase chain reaction (PCR). For example, SNP genotyping by MALDI-TOF mass spectrometry can performed using, for example, the Sequenom mass spectrometry system. In this method, multiplexed PCR is performed using more than one primer pairs each flanking different SNPs and then a minisequencing primer extension reaction performed in a single well using chain terminator nucleotides. The size of reaction products is determined directly by MALDI-TOF mass spectrometry, yielding the genotype information. Multiplexing permits determination of, for example, up to about 5 SNPs in a single well of a 384 well plate.

[020] Alternatively, fragment analysis for SNP detection can be performed on batches of several samples on a capillary electrophoresis system, for example an ABI PRISM® 3100 GENETIC ANALYZER (Applied Biosystems, Foster City, CA). For capillary electrophoretic analysis, the primers can be labeled using dyes, including, but not limited to FAM, HEX, NED, LIZ, ROX, TAMRA, PET and VIC.

[021] Single SNP allelic discrimination can further be carried out using the ABI PRISM® 7900HT Sequence Deatection System (Applied Biosystems, Foster City, CA),

which allows analysis of single nucleotide polymorphisms (SNPs) using the fluorogenic 5' nuclease assay.

[022] Yet another available method useful according to the present invention is an Arrayed Primer Extension (APEX) which is a resequencing method for rapid identification of polymorphisms that combines the efficiency of an microarray-based assay (alternative to gel-based methods, see, e.g. U.S. Patent No. 6,153,379 and Shumaker et al. Hum. Mutat. 7(4):346-354, 1996) with the Sanger nucleic acid sequencing method (Sanger et al., Proc. Natl. Acad. Sci. 74:5463-5467 (1977)). Generally, microarrays are microchips, for example glass slides, containing thousands of DNA segments in an ordered array, witch allows the simultaneous analysis of thousands of genetic markers.

[023] A yet another genotyping method useful according to the present invention is a solid-phase mini-sequencing technique, which is also based upon a primer extension reaction and can be used for genotyping of SNPs and can also be easily automated (U.S. Patent No. 6,013,431, Suomalainen et al. Mol. Biotechnol. Jun;15(2):123-31, 2000).

[024] In general, a single-base extension reaction is a modified cycle sequencing reaction in which only dideoxynucleotides (terminators) are present. When a terminator is incorporated onto a DNA strand, no further extension can occur on that strand. In a standard cycle sequencing reaction, terminators are present only in small concentrations along with high concentrations of typical nucleotides. In the single base extension reactions for SNP assays, two or more fluorescently labeled terminator nucleotides (corresponding to the two or more alleles present at the SNP to be typed) are used.

[025] **Haplotype analysis.** Haplotype is defined as a combination of alleles or single nucleotide polymorphisms (SNPs) of closely linked loci that are found in a single chromosome which tend to be inherited together. For a specific gene segment, there are often many theoretically possible combinations of SNPs, and therefore there are many theoretically possible haplotypes.

[026] Information about gene flow in a pedigree can be used to reconstruct likely haplotypes for families and individuals. In one embodiment, the present invention provides a method wherein a set of SNPs are first identified, each polymorphic SNP marker from a subject is then genotyped individually, for example using about 1-10 ng,

preferably about 5 ng genomic DNA. The genomic DNA sample is then diluted into about 1 copy of genomic template per dilution. The haplotype is determined by determining the SNP's in the diluted sample. For haplotype analysis, multiplex genotyping assays using, for example, primers which anneal directly before the SNP and can be extended using an extension primer and chain terminating nucleotides, are consequently carried out using about 3 pg or approximately 1 copy of genomic template, unless otherwise specified genomic DNA.

[027] Alternatively, the dilution of single copy of genomic template is performed directly and the SNPs are haplotyped without prior genotyping of the samples. In one embodiment, the genotyping is used as a quality control for the amplification reaction.

[028] The genomic region to be haplotyped using the method of the present invention is preferably at least about 8 kb, at least about 10 kb or more, at least about 15 kb or more, at least about 20 kb or more, at least about 25 kb or more, at least about 35 kb or more and at least about 40 kb or more. Most preferably the genomic region is about 25 kb ore more.

[029] In determining the haplotypes, the genotyping reactions are multiplexed which term is meant to include combining at least two, preferably more than three, four five or up to about 10 extension primers in the same reaction to identify, preferably more than 3, 4, 5, or up to at least 10 SNPs in the same genotyping reaction.

EXAMPLE

[030] We first investigated the effects of genomic DNA concentration on haplotyping efficiency. We used 3 picograms (pg), 5 pg and 9 pg (equivalent of 1, 1.6 and 3 genomic template copies) of genomic DNA for PCR amplification and genotyping of 3 SNPs in the CETP region from 12 individuals. Each 3-plex assay was repeated 12-18 times to evaluate the PCR and haplotyping efficiency. A typical assay result is summarized in Table 1. The copy number of the genomic DNA region of interest for very dilute DNA solutions is estimated by the Poisson distribution [13]. Haplotyping results are categorized into 4 groups (Table 1). Failed assays can result from either failed PCR amplification from single copy DNAs or simply no template present due to stochastic fluctuation of very dilute DNA solutions. Partially failed genotyping calls (or incomplete multiplexes) are those that have only 1 or 2 SNPs successfully genotyped.

This is most likely due to unsuccessful PCR for 1 or 2 of the SNP DNA regions, since in most cases the 3 SNP markers are present or absent at the same time due to the close proximity of the SNP markers (< 628 bp). Poisson distribution may also result in the presence both alleles in the solution and hence the inability to resolve the phase of the SNPs. Successful haplotyping analysis can be achieved when a single copy of the allele or multiple copies of the same allele are present and the genotyping is successful.

[031] Incomplete multiplex genotyping can be used to estimate the efficiency of genotyping from single copy DNA molecules. A partial genotyping call suggests the presence of the SNP DNA but a failure to genotype some of the SNPs. We typically observe 5-10% incomplete multiplex genotyping calls (Fig. 2), suggesting a PCR efficiency of 90-95% with single DNA molecules. This approach may overestimate the PCR efficiency, since we did not take the completely failed assays into account. We also carried out detailed comparison between observed and theoretical values of failed assays, successful haplotyping and the presence of both alleles (Fig. 2 and see methods section for details of calculation). Theoretical values are based on the Poisson distribution of very dilute DNA solutions and the assumption of 100% PCR amplification efficiency. The close agreement between theoretical estimate and experimental observation substantiates the earlier estimate of extremely high PCR efficiency with single DNA molecules. This high PCR efficiency is mainly due to the high efficiency of amplification of very short amplicons (typically 100 bp) and the high sensitivity of MALDI-TOF mass spectrometric detection of DNA oligonucleotides. This high PCR efficiency is absolutely crucial for high-throughput haplotyping analysis. With our current PCR efficiency, we can achieve 40-45% haplotyping efficiency with one single reaction using 3-4.5 pg genomic DNA. A replicate of 4 independent multiplex genotyping assays will enable about 90% of direct haplotyping efficiency.

[032] We next demonstrate an approach for determining haplotypes where there are too many markers to be determined in one multiplex genotyping assay. Overlapping informative SNPs can be used to combine haplotypes from several multiplex assays. Six SNP markers in an 8kb CETP genomic region were chosen, and 2 overlapping 4-plex genotyping assays were used for haplotyping analysis (Fig. 3). We were able to

determine the haplotypes of all 12 individuals for that genomic region, with absolutely no optimization of the assay system.

[033] The approach presented here provides a powerful and unique technology platform for direct molecular haplotyping analysis of long-range genomic regions. This approach is completely independent of pedigree genotype information. We have also incorporated this technique with the commercially available MassArray™ system for high-throughput applications. This technology can be extremely useful in large-scale haplotyping and haplotype-based diagnostics.

[034] **Genomic DNAs and oligo nucleotides.** Human genomic DNA samples used for haplotyping of the CETP locus were provided by SEQUENOM Inc. (San Diego, CA). These DNAs were isolated using the Puregene DNA isolation kit (Gentra Systems) from blood samples purchased from the Blood Bank (San Bernadino County, CA). The personal background of the blood donors is not accessible for these samples. Human genomic DNAs samples for haplotyping of a 25kb segment on chromosome 5q31 were CETP family DNAs purchased from Coriell Cell Repositories (see Table 3). Information on SNPs and oligonucleotides for genotyping is provided in Table 2.

[035] **Genotyping and haplotyping analysis.** Genotyping analyses were carried out using the MassArray™ system (SEQUENOM Inc.). Each SNP from every individual was first genotyped individually using 5 ng genomic DNA. For haplotyping analysis, multiplex genotyping assays were carried out using 3 pg (or approximately 1 copy of genomic template, unless otherwise specified) genomic DNA.

[036] **Analysis of effects of genomic DNA concentration on haplotyping.** To calculate the percentage of failed assays, we simply counted all failed assays (no calls for either SNP), divided by the total number of assays. We typically do 12 to 18 replicates for each 6 or 12 individuals. The percentage of incomplete assays is calculated in the same way. To calculate percentage of successful haplotyping and both alleles, we excluded the data from those individuals with homozygous haplotypes. Theoretical predictions are based on the Poisson distribution of very diluted DNA solutions, according to a published method [13].

Table 1 · Sample Haplotype analysis with triplex genotyping assay[a]

Repeat	Genotype Calls
1	GGC[b]
2	GGC
3	-[c]
4	-GC[d]
5	-
6	GGC
7	-
8	ACA
9	-GC
10	A/G C/G A/C[e]
11	ACA
12	ACA

[a]Genotypes of 3 SNP markers were determined with triplex assays from 3 pg genomic DNA.

[b]The 3 SNPs are G, G, C genotype respectively.

[c]Failed to genotype any of the 3 SNPs.

[d]Failed to genotype the first SNP, the rest two SNPs are G and C respectively.

[e]Failed to separate the two alleles, thus the genotypes are A/G, C/A and A/C for the 3 SNPs.

Table 2. Single nucleotide polymorphism (SNP) markers, their chromosomal locations, primer pairs to amplify the markers and terminator mixes used in the reaction.

SNP ID	Chrom. Position	PCR primer 1	PCR primer 2	Extension Primer	Term. Mix
rs289741	47282625	TCTACCAGCTTGGCTCCCTC	AAGTCCATCAGCAGCAGCAG	GGGAGTCAGCCCAGCTC	AC
rs289742	47282337	ACTGGTGAGACAATCCCTTC	CCACTGGCATTAAAGTGCTG	AGCCACAGAAGAAGGACTCC	AC
rs289744	47281997	TACCAGAAACCAGACCTCTG	AGTGCTGGACAGAAAGTGAG	TGAGGATGGTGGGAGGG	AC
rs289744[a]	47281997	TCTACCAGAAACCAGACCTC	AGTGCTGGACAGAAAGTGAG	ACCTCTGAGGGCCCCTTAC	CG
rs2228667[a]	47282820	CTCGAGTGATAATCTCAGGG	AGGTAGTGTTTACAGCCCTC	TGATGATGTCGAAGAGGCTCATG	CG
rs5882[a]	47284007	TTACGAGACATGACCTCAGG	GCATTTGATTGGCAGAGCAG	CTGCAGGAAGCTCTGGATG	CC
rs5882[b]	47284007	GCATTTGATTGGCAGAGCAG	TTACGAGACATGACCTCAGG	AGAGCAGCTCCGAGTCC	AC

rs5880[b]	47285008	GCAGCACATACTGGAAATCC	TTTCTCTCCCCAGGATATCG	GCTTTTTCTTAGAATAGGAGG	ACT
rs5881[a]	47288087	AGATCTTGGGCATCTTGAGG	ACCCCTGTCTTCCACAGGTT	TGGGCCTGGCTGGGGAAGC	CG
rs5881[b]	47288087	ACCCCTGTCTTCCACAGGTT	AGATCTTGGGCATCTTGAGG	TGTCTTCCACAGGTTGTCGGC	ACT
rs291044[a]	47288647	GTAAAACTGCAGCTGAGGAG	TGACTAGGTCAGGTCCCCTC	GGAGTATTTAAAGGAGAGACACACTAG	CG
rs291044[b]	47288647	TGACTAGGTCAGGTCCCCTC	GTAAAACTGCAGCTGAGGAG	CCCTCGTGCCACAGCCT	ACT
rs2033254[b]	47290114	GGACATCAAAGGAACAGGAC	ACTCACAATATTGGGCAGGC	CAAGGGGCTAAGGGAGAAG	ACT
IGR2198A_1[c]	506266[d]	GGGTTGCATGAGCATTAAGT	CACATCAAGGATAAGACTGC	ATCTCTTCAGTAGACGAAC	AC
IGR2175A_2	495082	TGGCCTTGATTCAAACCCTG	AGATGAAGGAAATCCCAAGG	TGCCACTAACATACATAGTAAC	AC
IGR2150A_1	482171	CCTTGGCTTGATAGTCAAAC	ATTTGGAGGAGTGCAGAGAG	AGTCAAACTCTCACCAC	AC

[a]Multiplex Group a
[b]Multiplex Group b
[c]SNP ID from ref
[d]Position of SNP from ref

Term. Mix = terminator nucleotide mix. Chrom. Position = chromosomal position

Table 3. DNA samples used in the Example.

Repository Number	Sample Type	Sample Description	Relation
GM12547	Lymphoblast	CEPH/FRENCH PEDIGREE 66	father
GM12548	Lymphoblast	CEPH/FRENCH PEDIGREE 66	mother
GM12549	Lymphoblast	CEPH/FRENCH PEDIGREE 66	son
GM12550	Lymphoblast	CEPH/FRENCH PEDIGREE 66	daughter
GM12551	Lymphoblast	CEPH/FRENCH PEDIGREE 66	daughter
GM12552	Lymphoblast	CEPH/FRENCH PEDIGREE 66	son
GM12553	Lymphoblast	CEPH/FRENCH PEDIGREE 66	daughter
GM12554	Lymphoblast	CEPH/FRENCH PEDIGREE 66	daughter
GM12555	Lymphoblast	CEPH/FRENCH PEDIGREE 66	son
GM12556	Lymphoblast	CEPH/FRENCH PEDIGREE 66	paternal grandfather
GM12557	Lymphoblast	CEPH/FRENCH PEDIGREE 66	paternal grandmother
GM12558	Lymphoblast	CEPH/FRENCH PEDIGREE 66	maternal grandfather
GM12559	Lymphoblast	CEPH/FRENCH PEDIGREE 66	maternal grandmother
GM07038	Lymphoblast	CEPH/UTAH PEDIGREE 1333	father
GM06987	Lymphoblast	CEPH/UTAH PEDIGREE 1333	mother
GM07004	Lymphoblast	CEPH/UTAH PEDIGREE 1333	son
GM07052	Lymphoblast	CEPH/UTAH PEDIGREE 1333	son
GM06982	Lymphoblast	CEPH/UTAH PEDIGREE 1333	son
GM07011	Lymphoblast	CEPH/UTAH PEDIGREE 1333	daughter
GM07009	Lymphoblast	CEPH/UTAH PEDIGREE 1333	son

GM07678	Lymphoblast CEPH/UTAH PEDIGREE 1333	son
GM07026	Lymphoblast CEPH/UTAH PEDIGREE 1333	son
GM07679	Lymphoblast CEPH/UTAH PEDIGREE 1333	son
GM07049	Lymphoblast CEPH/UTAH PEDIGREE 1333	paternal grandfather
GM07002	Lymphoblast CEPH/UTAH PEDIGREE 1333	paternal grandmother
GM07017	Lymphoblast CEPH/UTAH PEDIGREE 1333	maternal grandfather
GM07341	Lymphoblast CEPH/UTAH PEDIGREE 1333	maternal grandmother
GM11820	Lymphoblast CEPH/UTAH PEDIGREE 1333	daughter
GM07029	Lymphoblast CEPH/UTAH PEDIGREE 1340	father
GM07019	Lymphoblast CEPH/UTAH PEDIGREE 1340	mother
GM07062	Lymphoblast CEPH/UTAH PEDIGREE 1340	daughter
GM07053	Lymphoblast CEPH/UTAH PEDIGREE 1340	daughter
GM07008	Lymphoblast CEPH/UTAH PEDIGREE 1340	son
GM07040	Lymphoblast CEPH/UTAH PEDIGREE 1340	son
GM07342	Lymphoblast CEPH/UTAH PEDIGREE 1340	son
GM07027	Lymphoblast CEPH/UTAH PEDIGREE 1340	son
GM06994	Lymphoblast CEPH/UTAH PEDIGREE 1340	paternal grandfather
GM07000	Lymphoblast CEPH/UTAH PEDIGREE 1340	paternal grandmother
GM07022	Lymphoblast CEPH/UTAH PEDIGREE 1340	maternal grandfather
GM07056	Lymphoblast CEPH/UTAH PEDIGREE 1340	maternal grandmother
GM11821	Lymphoblast CEPH/UTAH PEDIGREE 1340	son
GM07349	Lymphoblast CEPH/UTAH PEDIGREE 1345	father
GM07348	Lymphoblast CEPH/UTAH PEDIGREE 1345	mother
GM07350	Lymphoblast CEPH/UTAH PEDIGREE 1345	daughter
GM07351	Lymphoblast CEPH/UTAH PEDIGREE 1345	son
GM07352	Lymphoblast CEPH/UTAH PEDIGREE 1345	son
GM07353	Lymphoblast CEPH/UTAH PEDIGREE 1345	son
GM07354	Lymphoblast CEPH/UTAH PEDIGREE 1345	daughter
GM07355	Lymphoblast CEPH/UTAH PEDIGREE 1345	son
GM07356	Lymphoblast CEPH/UTAH PEDIGREE 1345	son
GM07347	Lymphoblast CEPH/UTAH PEDIGREE 1345	paternal grandfather
GM07346	Lymphoblast CEPH/UTAH PEDIGREE 1345	paternal grandmother
GM07357	Lymphoblast CEPH/UTAH PEDIGREE 1345	maternal grandfather
GM07345	Lymphoblast CEPH/UTAH PEDIGREE 1345	maternal grandmother

REFERENCES

[037] The references cited herein and throughout the specification are herein incorporated by reference in their entirety.

1. Grupe, A. et al. In silico mapping of complex disease-related traits in mice. *Science* **292,** 1915-8. (2001).

2. Hirschhorn, J.N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet Med* **4**, 45-61. (2002).

3. Zhang, S., Pakstis, A.J., Kidd, K.K. & Zhao, H. Comparisons of two methods for haplotype reconstruction and haplotype frequency estimation from population data. *Am J Hum Genet* **69**, 906-14. (2001).

4. Templeton, A.R., Sing, C.F., Kessling, A. & Humphries, S. A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* **120**, 1145-54. (1988).

5. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44. (1999).

6. Judson, R., Stephens, J.C. & Windemuth, A. The predictive power of haplotypes in clinical response. *Pharmacogenomics* **1**, 15-26. (2000).

7. Martin, E.R. et al. Analysis of association at single nucleotide polymorphisms in the APOE region. *Genomics* **63**, 7-12. (2000).

8. Clark, A.G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* **7**, 111-22. (1990).

9. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89. (2001).

10. Ruano, G. & Kidd, K.K. Direct haplotyping of chromosomal segments from multiple heterozygotes via allele-specific PCR amplification. *Nucleic Acids Res* **17**, 8392. (1989).

11. Ruano, G., Kidd, K.K. & Stephens, J.C. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc Natl Acad Sci U S A* **87**, 6296-300. (1990).

12. Douglas, J.A., Boehnke, M., Gillanders, E., Trent, J.M. & Gruber, S.B. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat Genet* **28**, 361-4. (2001).

13. Stephens, J.C., Rogers, J. & Ruano, G. Theoretical underpinning of the single-molecule-dilution (SMD) method of direct haplotype resolution. *Am J Hum Genet* **46**, 1149-55. (1990).

14. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. High-resolution haplotype structure in the human genome. *Nat Genet* **29**, 229-32. (2001).

15. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9. (2002).

CLAIMS

We claim:

1. A method of haplotyping comprising the steps of identifying at least two SNPs within a genomic region; obtaining a DNA sample; amplifying the DNA sample comprising more than one copy of the genomic DNA with at least two primer pairs each capable of amplifying a different region flanking the at least two SNPs in separate reactions and individually genotyping the amplified DNA; diluting the DNA sample so that substantially only one copy of a genomic DNA is present in the dilution; amplifying the diluted DNA sample using the same primer pairs and performing a multiplex genotyping of the amplified DNA; after genotyping, determining the haplotype of the one copy of the genomic DNA from the single genomic DNA copy dilution.
2. A method of haplotyping comprising obtaining a genomic DNA sample; diluting the DNA sample so that substantially only one copy of a genomic DNA is present in the dilution; identifying at least two SNPs within a genomic region; amplifying the diluted DNA sample in a multiplex reaction using primer pairs designed to amplify the genomic regions flanking the SNPs; and determining the haplotype of the copy of the genomic DNA by multiplex genotyping.
3. The method of claim 2, wherein the genomic DNA sample is genotyped, in a separate reaction using more than one copy of the genomic DNA.
4. The method of any of the preceding claims, wherein the genotyping is performed using primer extension, terminator nucleotides and mass spectrometry.

FIGURE 1/3

A.

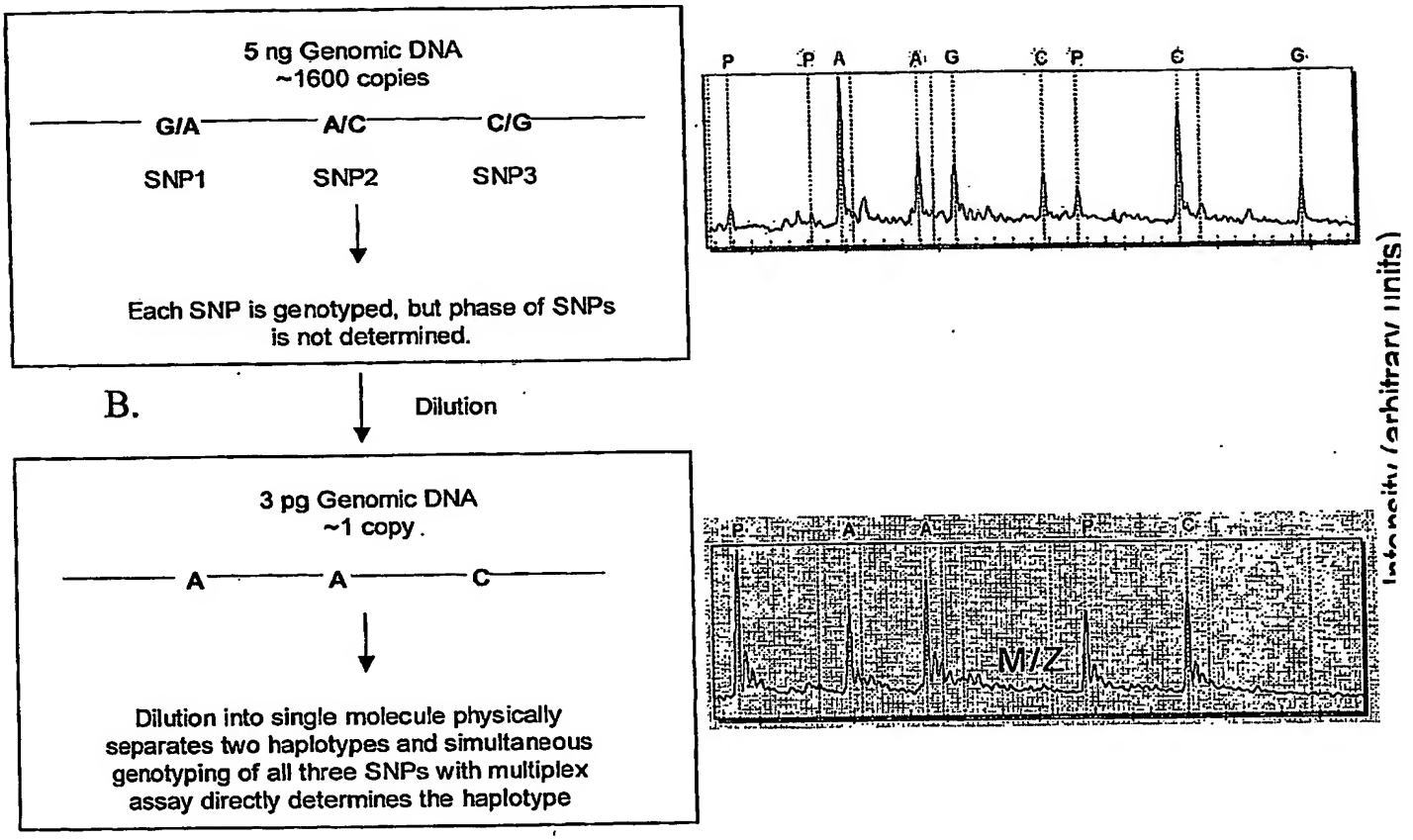

5 ng Genomic DNA
~1600 copies
G/A
A/C
C/G
SNP1
SNP2
SNP3
Each SNP is genotyped, but phase of SNPs is not determined.
B.
Dilution
3 pg Genomic DNA
~1 copy.
A
A
C
Dilution into single molecule physically separates two haplotypes and simultaneous genotyping of all three SNPs with multiplex assay directly determines the haplotype
P
P
A
A
G
C
P
C
G
M/Z
Intensity (arbitrary units)

FIGURE 2/3

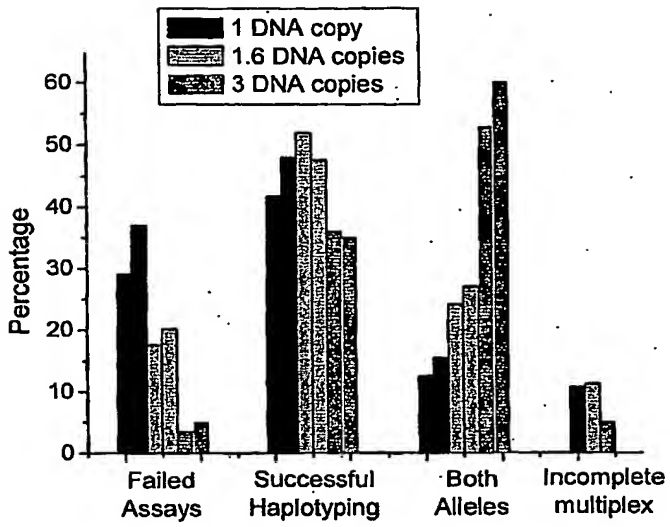
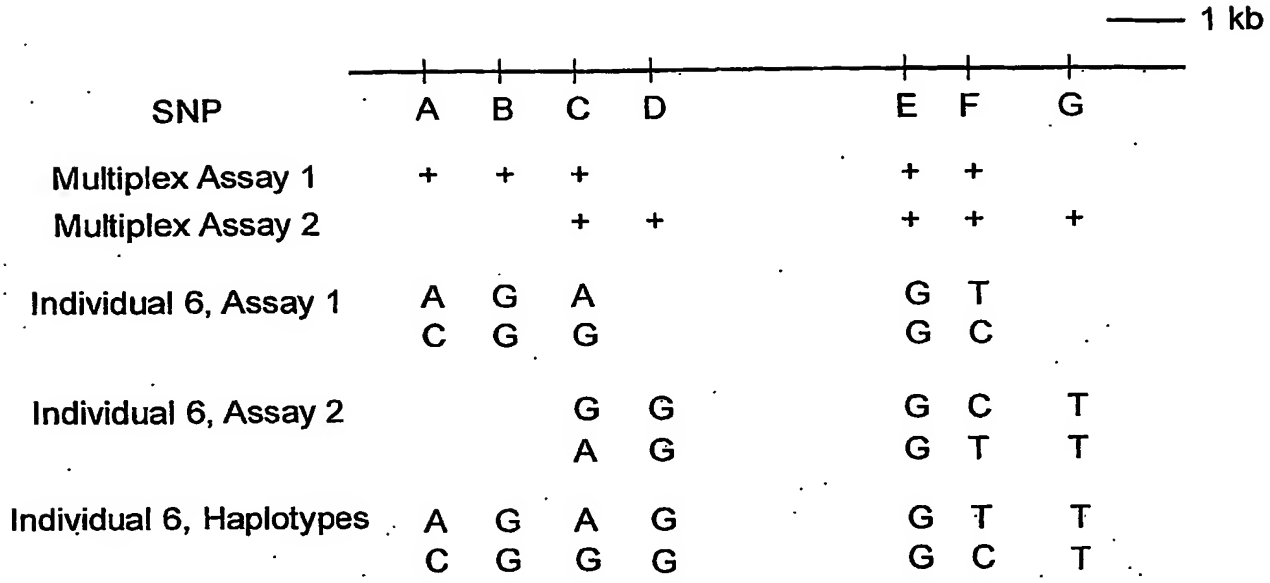

1 DNA copy
1.6 DNA copies
3 DNA copies
Percentage
0
10
20
30
40
50
60
Failed Assays
Successful Haplotyping
Both Alleles
Incomplete multiplex

FIGURE 3/3


1 kb
SNP
A
B
C
D
E
F
G
Multiplex Assay 1
Multiplex Assay 2
Individual 6, Assay 1
Individual 6, Assay 2
Individual 6, Haplotypes